frontiers

Probabilistic Power Flow Calculation Using Principal Component Analysis-Based Compressive Sensing

Tonghe Wang [1,2], **Hong Liang** [3] , **Junwei Cao**[4] **and Yuming Zhao** [5,*]

[1] *Guangzhou Institute of Energy Conversion, Chinese Academy of Sciences, Guangzhou, Guangdong, P.R. China*
[2] *Jiangsu Collaborative Innovation Center of Photovoltaic Science and Engineering, Changzhou University, Changzhou, P.R. China*
[3] *Meihua Holdings Group Co., Ltd., Langfang, Hebei, P.R. China*
[4] *Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, P.R. China*
[5] *School of Computer Science and Software, Zhaoqing University, Zhaoqing, Guangdong, P.R. China*

Correspondence*:
School of Computer Science and Software, Zhaoqing University, Zhaoqing, Guangdong, 526061, P.R. China
ymzhao@zqu.edu.cn

ABSTRACT

The increasing scale of the injection of renewable energy has brought about great uncertainty to the operation of power grid. In this situation, probabilistic power flow (PPF) calculation has been introduced to mitigate the low accuracy of traditional deterministic power flow calculation in describing the operation status and power flow distribution of power systems. Polynomial chaotic expansion (PCE) method has become popular in PPF analysis due to its high efficiency and accuracy, and sparse PCE has increased its capability of tackling the issue of dimension disaster. In this paper, we propose a principal component analysis-based compressive sensing (PCA-CS) algorithm solve the PPF problem. The l_1-optimization of CS is used to tackle the dimension disaster of sparse PCE, and PCA is included to further increase the sparsity of expansion coefficient matrix. Theoretical and numerical simulation results show that the proposed method can effectively improve the efficiency of PPF calculation in the case of random inputs with higher dimensions.

Keywords: probabilistic power flow, principal component analysis, compressive sensing, renewable energy

1 INTRODUCTION

In face of the global energy crisis and environmental pollution issues, countries are vigorously promoting the development and utilization of clean and renewable energy. However, the power of renewable energy generation devices, such as photovoltaics (PV) and wind turbines (WT), is usually affected by many uncertain factors, showing strong randomness and intermittence (Hua et al., 2021). With the increasing scale of PVs and WTs connected to the power grid, the uncertainty encountered by power systems will also

increase Hua et al. (2022b). As a result, the deterministic power flow analysis methods used to determine the operation status of interconnected power networks will no longer be applicable, and probabilistic power flow (PPF) emerge as the times require (Dalton et al., 2021). By effectively considering various uncertain factors of the energy system, PPF calculation is of great significance to help with system fault diagnosis, maintain system stability, and ensure system safety.

The essence of PPF is to obtain the statistical characteristics satisfied by the state variables of a power system by solving stochastic equations with random variables input. At present, mature methods of solving PPF equations mainly include simulation method, analytical method, approximation method, and polynomial chaotic expansion (PCE) method. Seeing the fact that analytical method and approximation method have difficulty in ensuring the calculation accuracy when the fluctuation of random inputs is large (Liang et al., 2021), we only provide details of simulation method and PCE method here:

- **Simulation method** stands for Monte Carlo (MC) method, as well as its improved versions (Constante-Flores and Illindala, 2019). It works as follows. First, a fairly large sample of input variables is obtained. Then, the corresponding solution is calculated based on the sample. Finally, subsequent statistical analysis on the solution is performed. Although this method has simple principle and convenient operation, it has low computational efficiency due to large amount of sampling, slow convergence speed and low computational efficiency (Liang et al., 2022).
- Recently, **PCE method** has been extensively adopted to solve PPF due to its great importance in uncertain quantization theory (Shen et al., 2020). Its first step is to expand random variables under a set of standard orthogonal basis made up by random polynomial functions. It then obtain the expansion coefficients and random variables by solving equations. The efficiency and accuracy of PCE method are relatively higher compared to other methods, but it needs to solve large-scale equations. Therefore, it is affected by the dimension disaster and is not satisfactory when solving problems with higher dimensions.

Sparse PCE is a common method to overcome the dimension disaster of traditional PCE method by reducing the number of bases in the polynomial expansion. In the calculation of probabilistic load flow, (Ma et al., 2021) proposes a sparse PCE method that reduces the number of basis functions of polynomial expansion by only preserving significant polynomial bases. To reduce the computational complexity of wideband configuration for periodic-grating wideband filters applied in optical devices under uncertain conditions, Papadopoulos et al. (2019) develops a sparse PCE method based on orthogonal matching pursuit. The principle of the proposed method is derived from **compressive sensing (CS)**, which is very popular in the field of signal and image processing (Blanco-Solano et al., 2021). When the PCE coefficients are sparse, combined with the principle of compressive sensing, the expression of random state variables by constructing the information matrix can be accurately reconstructed with the number of samples far lower than that of MC method. A similar idea is adopted in (Sun et al., 2019) when analyzing system sensitivity by solving PPF equations. Although PCE method is formally applied, PCE coefficients are solved by sampling to avoid solving large complicated equations, thus mitigating the issue of dimension disaster. At the same time, this method significantly reduces the sample size and enhances the operation efficiency compared with the traditional MC method.

Principal component analysis (PCA) method is widely used in the field of data dimensionality reduction by extracting the main features of data. By retaining only the principal components that contain most of the variance of the original data, PCA preserves the important information in the original data and reduces the components of redundancy or noise (Jaramillo et al., 2020). There have been some existing works that

use PCA in PPF analysis. To deal with the limitation of PPF calculation using traditional point estimation method, Li et al. (2020) uses PCA to reduce the correlation of original input random variables and improve the efficiency of the point estimate method. Considering the uncertainty in users' consumption behavior, the PPF analysis in (Memon et al., 2020) combines PCA method and PCE method to compress the number of output variables in the surrogate models of the least-squares support vector machine. In addition, the PPF calculation in (Le et al., 2021) also uses PCA to reduce the dimensionality of the dataset before it is partitioned into clusters by particle swarm optimization.

This paper extends the work of our previous conference paper (Liang et al., 2021). In a power grid with loads, PVs, and WTs that causes power flow fluctuation, this paper proposes a PPF calculation algorithm called Principal Component Analysis-based Compressive Sensing (PCA-CS). First, PCA-CS integrates the PCA theory in the decomposition of the covariance matrix of random state variables, and obtains a set of standard orthogonal bases composed of eigenvectors. Then, it transforms the PCE coefficients of random state variables to obtain a sparser expansion coefficient matrix. Moreover, we test our PCA-CS method on IEEE 118-bus system to show that PCA-CS has significantly improved the solution accuracy and computational efficiency compared with the original CS method and the traditional MC method. Compared with existing works, this paper adopt the idea of CS that reduces the number of expansion basis functions via l_1-optimization to overcome the dimensionality issue of PCE method. In addition, this paper theoretically proves that the expansion coefficient matrix is sparser after performing PCA-CS. According to the error theory of compressive sensing (Candes and Wakin, 2008), the sparser the expansion coefficient matrix is, the fewer samples are required to restore to the same accuracy. Therefore, our PCA-CS method can largely enhance the efficiency of PPF calculation while preserving the accuracy.

The remainder of this paper is arranged as follows: Section 2 explains in detail the PPF model of the power grid system we look into; Section 3 first introduces the traditional CS algorithm and then proposes the improved version, PCA-CS, with a detailed algorithm workflow; Section 4 provides numerical results and correlation analysis by simulation; Finally, Section 5 summarizes this paper and describes our future work plans.

2 MODEL OF PROBABILISTIC POWER FLOW

In this paper, we consider a power grid systems containing load nodes, PV nodes, and WT nodes, which are the main origins of the fluctuation in the power flow. This section describes the probabilistic models of loads, PVs, WTs, and the corresponding PPF equations.

2.1 Load Power Fluctuation Model

Due to many unpredictable factors such as geographical environment, time, and user behavior, the power fluctuation of load nodes could bring randomness to the power grid. This randomness is usually described by normal distribution, and its probability density function is as follows (Liu et al., 2017):

$$\begin{cases} f\left(P_L\right) = \frac{1}{\sqrt{2\pi}\sigma_{P_L}} \exp\left\{-\frac{\left(P_L-\mu_{P_L}\right)^2}{2\sigma_{P_L}^2}\right\}, \\ f\left(Q_L\right) = \frac{1}{\sqrt{2\pi}\sigma_{Q_L}} \exp\left\{-\frac{\left(Q_L-\mu_{Q_L}\right)^2}{2\sigma_{Q_L}^2}\right\}, \end{cases} \tag{1}$$

where:

- P_L is the load active power,

- μ_{P_L} is the expectation of P_L,
- σ_{P_L} is the stadard deviation of P_L,
- Q_L is the load reactive power,
- μ_{Q_L} is the expectation of Q_L,
- σ_{Q_L} is the standard deviation of Q_L.

2.2 Photovoltaic Power Fluctuation Model

PV power mainly depends on the intensity of solar irradiation, and its distribution roughly follows the beta distribution. Therefore, we use the beta distribution to describe the fluctuation of PV power, and the corresponding probability density function is (Rawat and Vadhera, 2018):

$$f\left(r_{PV}\right)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}r_{PV}^{\alpha-1}\left(1-r_{PV}\right)^{\beta-1}, \tag{2}$$

where:

- r_{PV} is the ratio of P_{PV}, the PV active output, to $P_{PV}^{\max}$, the maximum active power of PV power,
- $\Gamma(\cdot)$ is the gamma function,
- α and β are parameters of the beta distribution.

Generally speaking, PVs are connected to the power grid by means of constant power factor control. In other words, if we use φ_{PV} to represent the PV phase angle, then the power factor of PV $\cos\varphi_{PV}$ can be regarded as a constant. Then Q_{PV}, the PV reactive power, is therefore calculated by:

$$Q_{PV}=P_{PV}\tan\varphi_{PV}. \tag{3}$$

2.3 Wind Turbine Power Fluctuation Model

Similar to PV power, the randomness of WT power is mainly affected by wind speed, which is usually described by 2-parameter Weibull distribution. The corresponding probability density function is as follows (Gugliani et al., 2021):

$$f(v)=\frac{k}{c}\left(\frac{v}{c}\right)^{k-1}\exp\left\{-\left(\frac{v}{c}\right)\right\}, \tag{4}$$

where:

- v is the wind speed,
- k is the shape parameter of Weibull distribution,
- c is the scale parameter of Weibull distribution.

According to the technical principle of WT power generation, when the wind speed is less than a certain value or greater than a certain threshold, the WT power generation is zero. Therefore, the WT can work normally only within a reasonable wind speed range. Specifically, the active power and wind speed of wind

power generation meet the following piecewise function relationship:

$$P_{WT}(v) = \begin{cases} 0 & v < v_{ci} \\ \frac{P_r(v-v_{ci})}{v_r - v_{ci}} & v_{ci} \le v \le v_r \\ P_r & v_r \le v \le v_{co} \\ 0 & v \ge v_{co} \end{cases} \tag{5}$$

where

- $P_{WT}(v)$ is the active power when the wind speed is v,
- P_r is the rated power of the WT,
- v_{ci} is the cut-in wind speed,
- v_r is the rated wind speed,
- v_{co} is the cut-out wind speed.

By substituting (5) into (4), the probability distribution of P_{WT} can be obtained by:

$$f\left(P_{WT}\right) = \begin{cases} 1 - \exp\left\{-\left(\frac{v_{ci}}{c}\right)^k\right\} + \exp\left\{-\left(\frac{v_{co}}{c}\right)^k\right\} & P_{WT} = 0 \\ \frac{k}{k_1 c}\left(\frac{P_{WT}-k_2}{k_1 c}\right)^{k-1} \exp\left\{-\left(\frac{P_{WT}-k_2}{k_1 c}\right)^k\right\} & 0 \le P_{WT} \le P_r \\ \exp\left\{-\left(\frac{v_r}{c}\right)^k\right\} + \exp\left\{-\left(\frac{v_{co}}{c}\right)^k\right\} & P_{WT} = P_r \end{cases} \tag{6}$$

where k_1 and k_2 can be calculated as follows:

$$k_1 = \frac{P_r}{v_r - v_{ci}} \tag{7}$$

$$k_2 = -k_1 v_{ci}. \tag{8}$$

Similar to PV power generation, the reactive power of WT power generation Q_{WT} can be obtained by:

$$Q_{WT} = P_{WT} \tan \varphi_{WT}, \tag{9}$$

where φ_{WT} represents the phase angle of WT power generation. Since WTs are also connected to the power grid by constant power factor control, the power factor of WT power generation $\cos \varphi_{WT}$ is a constant as well.

2.4 Probabilistic Power Flow Model

Maintaining power balance at all times is fundamental for the normal operation of the power grid, which requires the power of each node i ($= 1, 2, \ldots, m$, where m represents the total number of grid nodes) to

meet the following power flow equation (Li et al., 2020):

$$\begin{cases} P_i = V_i \sum_{j=1}^{m} V_j \left(G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij}\right) \\ Q_i = V_i \sum_{j=1}^{m} V_j \left(G_{ij} \sin \delta_{ij} + B_{ij} \cos \delta_{ij}\right) \end{cases} \tag{10}$$

where:

- i can be a load, PV, or WT node,
- P_i is the active power of node i,
- Q_i is the reactive power of node i,
- V_i is the voltage amplitude of node i
- δ_{ij} represents the phase difference between node i and adjacent node j,
- G_{ij} and B_{ij} are parameters related to the transmission line that connects node i and node j.

Since the injected power of system nodes includes the randomness caused by loads, PV power generation, and WT power generation, P_i and Q_i are also regarded random variables. Therefore, (10) is the required PPF equation. By specifying the active and the reactive power of the node, it is possible to obtain the statistics of the node voltage amplitude and the phase angle, and then the corresponding line power flow is accordingly obtained.

For the sake of convenience, (10) will be condensed as follows:

$$\boldsymbol{X} = F(\boldsymbol{Y}), \tag{11}$$

where

- $\boldsymbol{X}$ consists of P_i and Q_i for all i,
- $\boldsymbol{Y}$ contains V_i and δ_{ij} for all i and j,
- $F(\cdot)$ is the function that maps $\boldsymbol{X}$ to $\boldsymbol{Y}$ determined by (10).

3 PRINCIPAL COMPONENT ANALYSIS-BASED COMPRESSIVE SENSING

The adoption of compressive sensing in PPF analysis aims to expand the random state variables by PCE, and then compress and restore them by using the sparsity of expansion coefficients. The PCA-CS algorithm proposed in this paper is to obtain a set of eigen-basis functions by decomposing the covariance of random variables before compression reduction, so as to make the random variables sparser under the expression of this set of basis functions, and then compress and restore under this set of basis functions. According to the error theory of compressive sensing, improving the sparsity can reduce the number of sample solutions required for restoration, and then improve the solution efficiency.

To sum up, this section will describe the proposed algorithm in detail from the following three parts:

1. Traditional PCE method of random variables in probability space;
2. The covariance matrix of random variables is decomposed to obtain the eigen-basis functions;

3. The sparse expression of random variables is obtained by compression reduction under the expansion of eigen-basis functions.

3.1 Traditional PCE Method

Assume that the input variable $\boldsymbol{Y} = [y_1, y_2, \ldots, y_d]$, the power distribution of all nodes, is a d-dimensional variable with all entries following independent normal distribution. Otherwise, we can mitigate the correlation by Nataf transform (see (Lin et al., 2020) for more details). Then, we expand (11) with Hermite orthogonal polynomials:

$$\boldsymbol{X} \approx \sum_{|\boldsymbol{I}| \leq p} s_{\boldsymbol{I}} \psi_{\boldsymbol{I}}(\boldsymbol{Y}), \tag{12}$$

where:

- multilevel index $\boldsymbol{I} = (i_1, i_2, \ldots, i_d) \in \mathbb{N}_0^d$ satisfies $|\boldsymbol{I}| = i_1 + i_2 + \ldots + i_d$,
- p is the expansion order of polynomials,
- $\psi_{\boldsymbol{I}}$ is a multivariable orthogonal Hermite basis function calculated by:

$$\psi_{\boldsymbol{I}}(\boldsymbol{Y}) = h_{i_1}(y_1) h_{i_2}(y_2) \ldots h_{i_d}(y_d), \tag{13}$$

- $h_i(\cdot)$ is a univariate Hermite functions.

Orthogonal polynomial $\psi_{\boldsymbol{I}}$ satisfies:

$$\mathbb{E}\left[\psi_{\boldsymbol{I}}(\boldsymbol{Y}) \psi_{\boldsymbol{J}}(\boldsymbol{Y})\right] = \int \psi_{\boldsymbol{I}}(\boldsymbol{Y}) \psi_{\boldsymbol{J}}(\boldsymbol{Y}) \rho(\boldsymbol{Y}) \mathrm{d}\boldsymbol{Y} = \gamma_{\boldsymbol{I}} \chi_{\boldsymbol{I}\boldsymbol{J}}, \tag{14}$$

where:

- $\chi_{\boldsymbol{I}\boldsymbol{J}}$ is the Kronecker function (i.e., 1 at that time $\boldsymbol{I} = \boldsymbol{J}$ and 0 in other cases),
- $\rho(\boldsymbol{Y})$ is the joint probability density of $\boldsymbol{Y}$,
- $\gamma_{\boldsymbol{I}}$ is a constant.

The goal is to obtain the expansion coefficient $s_{\boldsymbol{I}}$ in (12), and the number of terms in $s_{\boldsymbol{I}}$ is:

$$P = \frac{(p+d)!}{p!d!}. \tag{15}$$

According to the Askey scheme, Hermite orthogonal basis functions are efficient in approximating independent normal random variables (Son and Du, 2021). Next, the moment and probability distributions of random variables are calculated by obtaining the expansion coefficients.

3.2 PCA for Extracting Eigen-Basis Functions

At present, the existing methods for solving the coefficients of PCE, such as Garrerkin projection method (Wu et al., 2017) and collocation point method (Tang et al., 2016), will face problems such as complicated operation and dimension disaster. Because the expansion of random variables is usually sparse, compressive sensing algorithm can be used to restore the expansion coefficients. By introducing PCA, PCA-CS further improves the application of traditional compressive sensing algorithm in PPF solution. It

overcomes the dimensional problem of the PCE method while inheriting the feature that the MC method is easy to operate.

In more detail, PCA method is used to map random variables to their eigen-space, so that the expansion of random variables is sparser in the expression of eigen-space.

First, the eigen-basis function $\{\phi_j\}_{j=1}^m$ is obtained by solving the following eigen-decomposition problem:

$$\boldsymbol{C}(\boldsymbol{X}, \boldsymbol{X}^{\mathrm{T}})\phi_j = \mu_j \phi_j, \tag{16}$$

where:

- $\boldsymbol{C}(\boldsymbol{X}, \boldsymbol{X}^{\mathrm{T}})$ represents the covariance matrix of random variable $\boldsymbol{X}$ and $\boldsymbol{X}^{\mathrm{T}}$ (the transpose of $\boldsymbol{X}$) that can be calculated as follows:

$$\boldsymbol{C}(\boldsymbol{X}, \boldsymbol{X}^{\mathrm{T}}) = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})\left(\boldsymbol{X}^{\mathrm{T}} - \mathbb{E}\boldsymbol{X}^{\mathrm{T}}\right)\right]; \tag{17}$$

- $\{\mu_j\}_{j=1}^m$ represents the eigenvalues corresponding to the eigenvectors $\{\phi_j\}_{j=1}^m$.

During calculation, the eigenvalues are arranged in descending order by default. Because the covariance matrix is symmetric and positive definite, the eigenvalues $\{\mu_j\}_{j=1}^m$ are nonnegative real numbers.

After obtaining the eigen-basis function $\{\phi_j\}_{j=1}^m$, expand the random variable $\boldsymbol{X}$ under this set of basis functions, then:

$$\boldsymbol{X} \approx \sum_{i=1}^{P} s_i \psi_i \approx \sum_{k=1}^{P}\sum_{j=1}^{m} S_{ij}\psi_i\phi_j. \tag{18}$$

3.3 Improved Compressive Sensing Algorithm

After the expansion of $\boldsymbol{X}$ is obtained, the expansion coefficient $\boldsymbol{S} = (S_{ij})_{P\times m}$ of $\boldsymbol{X}$ can be restored through the sparse restoration algorithm of CS.

3.3.1 Solution Flow of Traditional Compressive Sensing Algorithm

Since the coefficient matrix $\boldsymbol{S}$ in expansion (12) is sparse, according to the sparse reduction theory (Marques et al., 2019), the expansion coefficient $\boldsymbol{S}$ can be restored by sampling. The specific process is as follows: select H random sample points $[\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \ldots, \boldsymbol{Y}^{(H)}]$ of $\boldsymbol{Y}$, and then bring them into PPF equation (11) respectively to obtain corresponding sample solutions $\boldsymbol{u} = [\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(H)}]^{\mathrm{T}}$, which meet the equation:

$$\boldsymbol{u} = \boldsymbol{\Psi}\bar{\boldsymbol{s}}, \tag{19}$$

where $\boldsymbol{\Psi}$ is called the measurement matrix, which is obtained by bringing multiple Hermite orthogonal polynomials from random sample points $[\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \ldots, \boldsymbol{Y}^{(H)}]$, namely:

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1(\boldsymbol{Y}^{(1)}) & \psi_2(\boldsymbol{Y}^{(1)}) & \cdots & \psi_P(\boldsymbol{Y}^{(1)}) \\ \psi_1(\boldsymbol{Y}^{(2)}) & \psi_2(\boldsymbol{Y}^{(2)}) & \cdots & \psi_P(\boldsymbol{Y}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\boldsymbol{Y}^{(H)}) & \psi_2(\boldsymbol{Y}^{(H)}) & \cdots & \psi_P(\boldsymbol{Y}^{(H)}) \end{bmatrix}, \tag{20}$$

and $\bar{\boldsymbol{s}} = [s_1, s_2, \ldots, s_P]^{\mathrm{T}}$.

The dimension of $\boldsymbol{\Psi}$ is $H \times P$. When $P > H$, the system of (19) is an underdetermined system of equations, and the coefficients $\bar{\boldsymbol{s}}$ have infinite solutions. Additional constraints need to be imposed in order to guarantee the uniqueness of $\bar{\boldsymbol{s}}$. It is usually hoped to reduce the number of expansion terms (and therefore increase the coefficient sparsity). Therefore, the constraint on increasing sparsity becomes:

$$\min \|\bar{\boldsymbol{s}}\|_0 \quad \text{s.t.} \quad \boldsymbol{\Psi}\bar{\boldsymbol{s}} = \boldsymbol{u}, \tag{21}$$

where $\|\cdot\|_0$ refers to the l_0-norm, which is calculated as the element count in $\bar{\boldsymbol{s}}$ that are not zero.

Seeing the fact that the l_0-norm is discontinuous, optimization problem (21) is NP hard to solve, so it needs to be relaxed. Among them, l_1-norm represents the sum of the absolute values of all elements in the matrix and is the most commonly used relaxation method. Thus, problem (21) becomes :

$$\min \|\bar{\boldsymbol{s}}\|_1 \quad \text{s.t.} \quad \boldsymbol{\Psi}\bar{\boldsymbol{s}} = \boldsymbol{u}, \tag{22}$$

which is also called an l_1-optimization problem.

It has been shown that the solution of problem (21) can be accurately approximated by solving problem (22), and the solution of l_1-optimization problem is more convenient (Wang et al., 2020). It can be solved by the orthogonal matching pursuit (OMP) algorithm (Papadopoulos et al., 2019).

3.3.2 PCA-CS

The proposed PCA-CS algorithm is based on the traditional CS algorithm. Through the PCA theory, the random variables are transformed to make their expression coefficients sparser, and then compressed and restored. According to (18), sample solutions $\boldsymbol{u} = [\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(H)}]^{\mathrm{T}}$ are written by:

$$\boldsymbol{u} \approx \boldsymbol{\Psi}\boldsymbol{S}\boldsymbol{\Phi}, \tag{23}$$

where:

- $\boldsymbol{\Psi}$ is the measurement matrix in the same form as (20),
- $\boldsymbol{\Phi}$ is the eigen-matrix composed of eigen-basis function,
- $\boldsymbol{S}$ is the expansion coefficient to be solved.

Furthermore, the coefficient $\boldsymbol{S}$ can be reduced through the following l_1-optimization problem:

$$\min \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \boldsymbol{\Psi}\boldsymbol{S}\boldsymbol{\Phi} = \boldsymbol{u}. \tag{24}$$

The sparsity of $\boldsymbol{S}$ is given by the following theorem:

THEOREM 1. *The expansion coefficient $\boldsymbol{S}$ of $\boldsymbol{X}$ and eigenvalues $\{\mu\}_{j=1}^{m}$ satisfy the following relationship:*

$$\sum_{i=2}^{P} |S_{ij}|^2 \approx \mu_j, \quad j = 1, 2, \ldots, m, \tag{25}$$

where μ_j represents the j-th eigenvalue of the eigen-decomposition problem.

PROOF OF THEOREM 1. Hermite polynomial ψ_i satisfies

$$\mathbb{E}(\psi_i) = \begin{cases} 1 & i = 1 \\ 0 & i \neq 1 \end{cases} \tag{26}$$

so

$$\mathbb{E}\boldsymbol{X} \approx \mathbb{E}\left[\sum_{i=1}^{P}\sum_{j=1}^{m} S_{ij}\psi_i\phi_j\right] = \sum_{j=1}^{m} S_{1j}\phi_j. \tag{27}$$

Combining (27) and (17), we can get:

$$\begin{aligned} \boldsymbol{C}(\boldsymbol{X}, \boldsymbol{X}^{\mathrm{T}}) &= \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})\left(\boldsymbol{X}^{\mathrm{T}} - \mathbb{E}\boldsymbol{X}^{\mathrm{T}}\right)\right] \\ &= \left(\sum_{i=2}^{P}\sum_{j=1}^{m} S_{ij}\psi_i\phi_j\right)\left(\sum_{i=2}^{P}\sum_{j'=1}^{m} S_{ij'}\psi_i\phi_{j'}\right) \\ &\approx \sum_{i=2}^{P}\left(\sum_{j=1}^{m} S_{ij}\phi_j\right)\left(\sum_{j'=1}^{m} S_{ij'}\phi_{j'}\right) \end{aligned} \tag{28}$$

By introducing (28) into eigen-decomposition equation (16):

$$\mu_j\phi_j \approx \sum_{i=2}^{P}\left(\sum_{j=1}^{m} S_{ij}\phi_j\right)\left(\sum_{j'=1}^{m} S_{ij'}\phi_{j'}\right)\phi_j = \sum_{i=2}^{P}|S_{ij}|^2\phi_j. \tag{29}$$

Then the theorem is proved. □

Theorem 1 shows that the value of each column of the expanded coefficient matrix of random variables can be roughly controlled by the corresponding eigenvalue. In addition, in practical application, the eigenvalues in descending order decline very fast. Therefore, when the eigenvalues tend to 0, the component element values of the corresponding columns of the coefficient matrix tend to 0. In other words, the faster the eigenvalue decreases, the sparser the coefficient matrix is represented by this set of eigen-basis functions. This will also be proved in the following numerical examples.

To sum up, the whole PPF solution process is as follows:

1. Perform the PCE on input variable $\boldsymbol{X}$ according to (12);
2. Randomly select H sample points $[\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(H)}]$ based on normal distribution and obtain corresponding solutions $\boldsymbol{u} = [\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(H)}]^{\mathrm{T}}$ according to the Newton-Raphson method described in (Liu et al., 2020);
3. Solve the eigen-decomposition problem (16) and obtain eigen-basis functions $\{\phi_j\}_{j=1}^{m}$;
4. Under the expression of eigen-basis functions, construct the compression reduction problem (22);
5. Obtain the coefficient matrix $\boldsymbol{S}$ by OMP algorithm;
6. Bring the coefficient $\boldsymbol{S}$ back to the expansion (18) to obtain the statistics of $\boldsymbol{Y}$, the state variable, and then analyze it.

Table 1. Parameters of random loads.

Load	μ_{P_L} (MW)	μ_{Q_L} (MW)	σ_{P_L} (MW)	σ_{Q_L} (MW)
1	3.2	2.3	0.16	0.115
2	4.4	4.2	0.22	0.210
3	7.5	6.5	0.375	0.325
4	2.8	1.6	0.140	0.080
5	8.3	4.4	0.415	0.220
6	4.7	1.8	0.235	0.090
7	5.8	2.7	0.290	0.135
8	6.3	5.2	0.315	0.260

Table 2. Parameters of WTs.

WT	P_r (MW)	v_{ci}	v_r	v_{co}	c	k	$\cos\varphi_{WT}$
1	13	3	20	10	8	3	0.9
2	20	3	16	11	8	3	0.9
3	25	2.4	19	12.4	8	3	0.9
4	15	3.5	23	13.6	8	3	0.9
5	10	2	15	11.9	8	3	0.9
6	16	4	22	10.7	8	3	0.9

Table 3. Parameters of PVs.

PV	$P_{PV}^{\max}$ (MW)	α	β	$\cos\varphi_{PV}$
1	50	0.9	0.8	0.95
2	60	0.8	0.85	0.95
3	55	0.85	0.75	0.95
4	70	0.7	0.9	0.95
5	45	0.8	0.8	0.95
6	80	0.75	0.9	0.95

4 NUMERICAL ANALYSIS

In this section, the PCA-CS algorithm for PPF proposed in Section 3 is verified on IEEE 118 node. We use MATLAB r2020 as the develop tool of our program, and we perform PPF calculation on a laptop with Intel i7-8586u CPU with the help of Matpower package.

4.1 Simulation Settings

In this example, on the basic grid of IEEE 118 node, WTs are connected at nodes 6, 15, 42, 60, 92, 115 respectively, PVs are connected at nodes 2, 7, 27, 41, 58, 98, and nodes 3, 11, 23, 50, 57, 75, 84, 88, 93, 102 are selected as load nodes. The parameter settings of loads, WTs, and PVs are listed in Table 1, 2, and 3 respectively.

4.2 Simulation Results

In order to verify the effectiveness of PCA-CS algorithm in solving PPF equation, this section makes numerical analysis of PCA-CS from the aspects of algorithm accuracy and calculation efficiency, and compares it with MC method and traditional CS method. We use the results of traditional MC method with a sampling scale of 100,000 times as the error reference standard.

Table 4. Relative errors of three methods with 30 sample points.

Method	$V_{\bar{E}}$	$V_{\bar{\sigma}}$	$\delta_{\bar{E}}$	$\delta_{\bar{\sigma}}$
PCA-CS	7.23×10^{-8}	5.29×10^{-9}	2.95×10^{-4}	2.61×10^{-3}
CS	8.45×10^{-7}	3.93×10^{-9}	1.13×10^{-4}	2.08×10^{-3}
MC	6.42×10^{-3}	5.27×10^{-4}	1.76×10^{-2}	9.34×10^{-2}

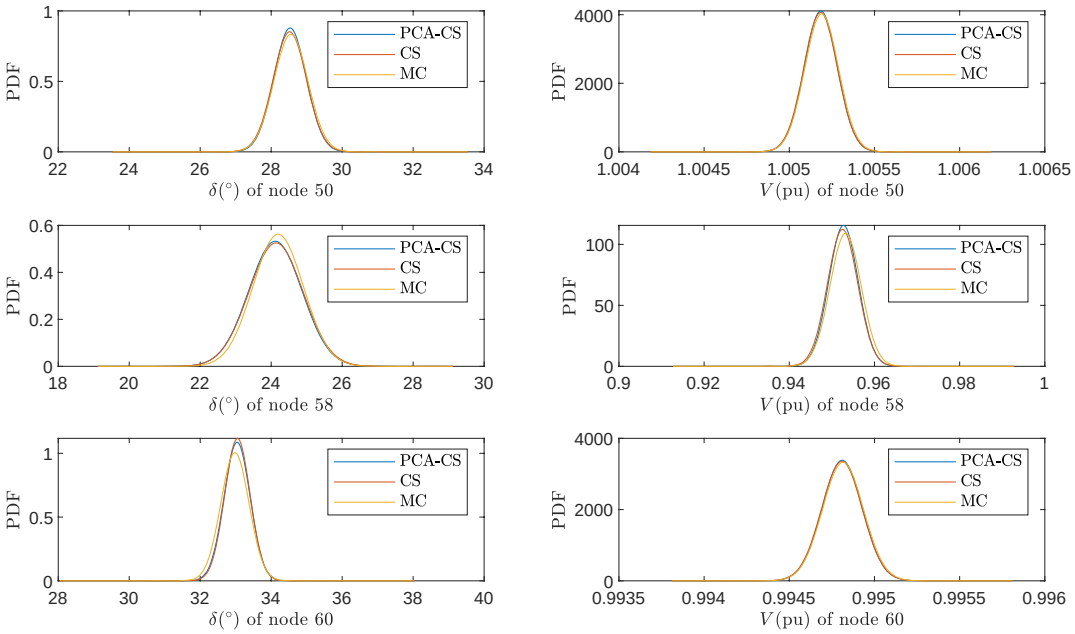

Figure 1. Probability density function (PDF) of node voltage amplitude (V) and phase angle δ.

4.2.1 Algorithm Accuracy

In order to evaluate the accuracy of PCA-CS, we need to select performance indicators to measure the accuracy first. Table 4 shows the relative error values of the mean (with subscription $\bar{E}$) and standard deviation (with subscription $\bar{\delta}$) of V and δ obtained by PCA-CS, CS, and MC under 30 sample points. We can see from the data in Table 4 that when the number of samples is relatively small, the output state variables of PPF obtained by PCA-CS and traditional CS algorithm have higher accuracy, while the accuracy of MC method is lower. This is because MC method is half-order convergent and requires a large number of sample solutions to achieve the corresponding accuracy, which is also the defect of MC algorithm.

Specifically, three representative random nodes 50, 58, and 60 are selected here to visualize the probability distribution of output state variables of random load, PV, and WT node respectively. Figure 1 shows the probability density functions of V and δ of nodes 50, 58, and 60 obtained by PCA-CS algorithm and CS algorithm, and compares them with the standard reference obtained by MC method with 100,000 sample solutions. It can be seen from the figure that the distribution functions calculated by PCA-CS and CS algorithm basically coincide with each other and are consistent with the standard reference solution, which further shows the accuracy of PCA-CS.

From the numerical analysis of accuracy, we can see that although PCA-CS performs a basis function transformation through principal component extraction based on the traditional CS method, it does not affect the accuracy of the results. This is because PCA-CS algorithm only obtains a set of orthogonal eigen-basis functions through the eigen-decomposition of PCA, but does not abandon the basis functions corresponding to small eigenvalues when expanding random variables, thus ensuring the accuracy of transformation to the greatest extent. On the contrary, after the transformation, the sparsity of the expansion is improved. According to the error theory of compressive sensing (Candes and Wakin, 2008), the sparsity is improved and the number of sample solutions required to achieve the same accuracy is reduced. Therefore, compared with the traditional CS method, PCA-CS can further improve the operation efficiency under the same accuracy.

4.2.2 Computational Efficiency

After verifying the accuracy, we further explains the advantages of PCA-CS algorithm in computational efficiency through experimental data. First, we introduce the concept of sparse ratio λ that can be calculated as follows:

$$\lambda = \frac{\#\{|\boldsymbol{S}| \geq \tau\}}{P \times 118} \tag{30}$$

where:

- $\#\{|\boldsymbol{S}| \geq \tau\}$ is the element counts in matrix $\boldsymbol{S}$ whose elements are greater than or equal to the threshold τ,
- p represents the dimension equal to the random basis function.

In other words, sparse ratio λ indicates the proportion of the coefficients in $\boldsymbol{S}$ that are greater than or equal to a reference value τ. The highest order of the truncated basis here is $p = 2$, and the dimension of the random variable $d = 40$. The number of basis is therefore calculated by:

$$P = \frac{(p+d)!}{p!d!} = \frac{(40+2)!}{40!2!} = 881.$$

The dimension of the expansion coefficient $\boldsymbol{S}$ is 881×118-dimensional. The smaller the sparse ratio λ is, the sparser the sparse matrix is.

Figure 2 and 3 respectively show the visual display of the sparsity of PCE coefficients of V and δ of 118 nodes calculated by PCA-CS and traditional CS algorithm and the value of total sparsity ratio.

From Figure 2 and 3, we can see the sparsity of V and δ of each node under the Hermite orthogonal expansion. The average sparsity of the two methods is less than 1%, which shows the applicability of compressed sensing idea. By comparison, after the introduction of principal component analysis, the sparsity of the expansion coefficient of voltage amplitude V decreases from 3.14% to 2.41%, the sparsity increases by 23.2%, the sparsity of phase angle δ decreases from 9.34% to 5.58%, and the sparsity increases by 40.3%. It is proved that the sparsity of the results of PCA-CS algorithm is significantly improved compared with CS algorithm, and the results accord with the analysis of Theorem 1. Therefore, PCA-CS algorithm can achieve the same calculation accuracy as CS algorithm through a smaller number of sample solutions. The error convergence comparison and calculation time comparison data of the two methods are given in Figure 4 and Table 5 respectively.

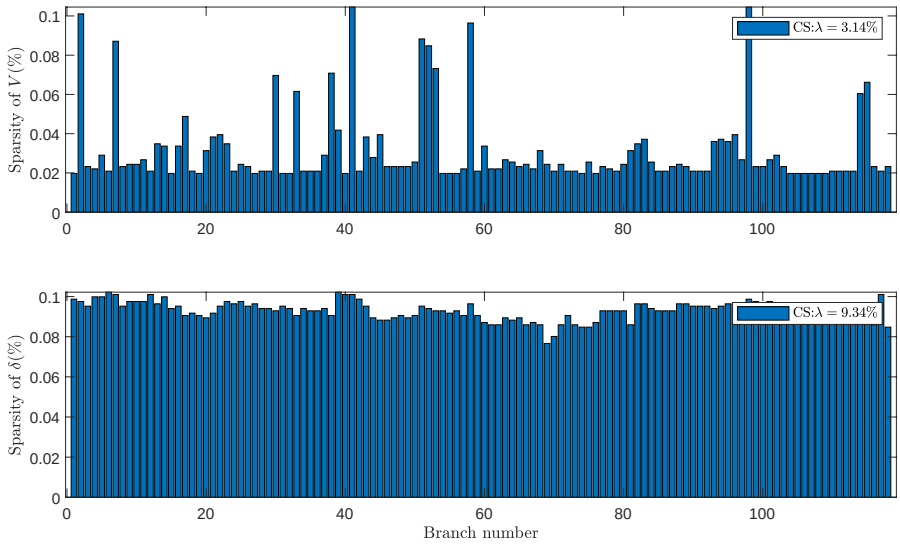

Figure 2. Sparsity of coefficient matrix of CS method.

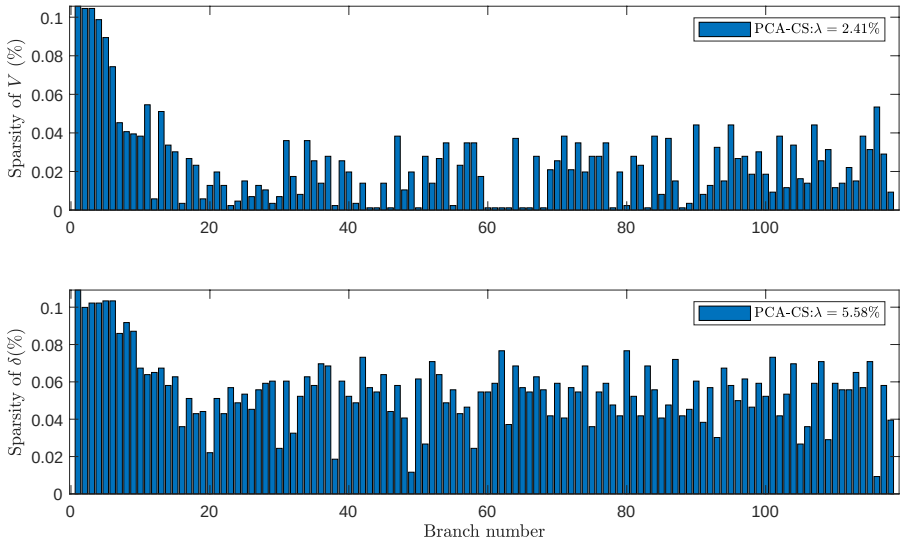

Figure 3. Sparsity of coefficient matrix of PCA-CS method.

Table 5. Comparison of computational efficiency of PCA-CS, CS, and MC methods.

Method	$V_{\bar{E}}$	Sample number	Computation time (s)
PCA-CS	7.85×10^{-7}	15	10.43
CS	7.93×10^{-7}	40	20.69
MC	8.42×10^{-7}	10000	4328.47

According to Figure 4, the errors of both PCA-CS and CS fall as the number of sample points increases at the beginning, and then the errors are basically unchanged after the sample scale reaches a certain level. However, at the same sample level, the error of PCA-CS algorithm is lower than that of CS algorithm.

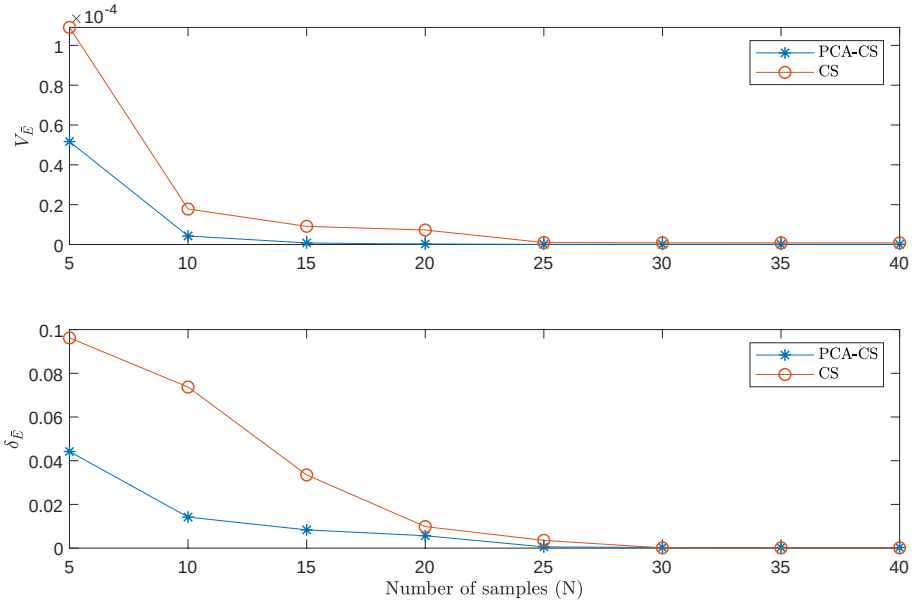

Figure 4. Error estimation of PCA-CS and CS.

According to the data in Table 5, to reach the same accuracy level, the running time of PCA-CS is roughly 10 s less than CS, and the time consumption of the two methods is much less than that of MC. Like MC, the PCA-CS uses sample solution in PPF calculation to simplify the implementation process of the algorithm.

To sum up, PCA-CS can significantly enhance the solution efficiency with little compromise in accuracy.

5 CONCLUSION

In this paper, we study the PPF calculation problem with the consideration of the input uncertainty caused by the injection of PVs and WTs into the power grid system. We propose a PCA-CS algorithm for PPF calculation uses the principle of PCA to perform eigen-transformation on the result of PCE expansion. This method is similar to the traditional CS method and MC method. It inherits the advantage of simple principle, easy implementation, and more suitable for practical application. In our future research, we will consider continuing to improve the PCA-CS algorithm by using the principles of weighted l_1-optimization in optimization theory to further enhancing the solution efficiency (Sun et al., 2016). Moreover, given the excellent performance of artificial intelligence (AI) in dealing with dimensional disasters, we can also consider solving PPF equations using AI-based methods (Hua et al., 2022a). Based on PPF analysis, we can further consider studying probabilistic stability control of microgrid systems (Hua et al., 2019).

CONFLICT OF INTEREST STATEMENT

Author Hong Liang is employed by Meihua Holdings Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Tonghe Wang: Conceptualization, Investigation, Writing-Original Draft; Hong Liang: Methodology, Software, Formal Analysis, Data Curation, Visualization; Junwei Cao: Supervision, Project administration, Funding acquisition; Yuming Zhao: Resources, Validation, Writing-Review & Editing.

FUNDING

This work is supported by the China National Key R&D Program intergovernmental international scientific and technological innovation cooperation project (2019YFE0120200).

REFERENCES

Blanco-Solano, J., Petit, J. F., Ordóñez-Plata, G., Kagan, N., and Almeida, C. F. M. (2021). Voltage sag state estimator based on compressive sensing in distribution systems. *International Journal of Electrical Power & Energy Systems* 130, 106892. doi:10.1016/j.ijepes.2021.106892

Candes, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine* 25, 21–30. doi:10.1109/msp.2007.914731

Constante-Flores, G. E. and Illindala, M. S. (2019). Data-driven probabilistic power flow analysis for a distribution system with renewable energy sources using monte carlo simulation. *IEEE Transactions on Industry Applications* 55, 174–181. doi:10.1109/tia.2018.2867332

Dalton, A., Bekker, B., and Koivisto, M. J. (2021). Classified atmospheric states as operating scenarios in probabilistic power flow analysis for networks with high levels of wind power. *Energy Reports* 7, 3775–3784. doi:10.1016/j.egyr.2021.06.060

Gugliani, G. K., Sarkar, A., Ley, C., and Matsagar, V. (2021). Identification of optimum wind turbine parameters for varying wind climates using a novel month-based turbine performance index. *Renewable Energy* 171, 902–914. doi:10.1016/j.renene.2021.02.141

Hua, H., Li, Y., Wang, T., Dong, N., Li, W., and Cao, J. (2022a). Edge computing with artificial intelligence: A machine learning perspective. *ACM Computing Surveys* doi:10.1145/3555802

Hua, H., Qin, Y., Hao, C., and Cao, J. (2019). Stochastic optimal control for energy internet: A bottom-up energy management approach. *IEEE Transactions on Industrial Informatics* 15, 1788–1797

Hua, H., Qin, Z., Dong, N., Qin, Y., Ye, M., Wang, Z., et al. (2022b). Data-driven dynamical control for bottom-up energy internet system. *IEEE Transactions on Sustainable Energy* 13, 315–327. doi:10.1109/tste.2021.3110294

Hua, H., Wei, Z., Qin, Y., Wang, T., Li, L., and Cao, J. (2021). Review of distributed control and optimization in energy internet: From traditional methods to artificial intelligence-based methods. *IET Cyber-Physical Systems: Theory & Applications* 6, 63–79. doi:10.1049/cps2.12007

Jaramillo, A. F. M., Laverty, D. M., del Rincon, J. M., Hastings, J., and Morrow, D. J. (2020). Supervised non-intrusive load monitoring algorithm for electric vehicle identification. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (IEEE). doi:10.1109/i2mtc43012.2020.9128529

Le, D. D., Ngo, D. V., Nguyen, N. T. A., and Huynh, K. V. (2021). Probabilistic assessment of power systems with renewable energy sources based on an improved analytical approach. *International Journal of Renewable Energy Development* 10, 811–818. doi:10.14710/ijred.2021.38226

Li, H., Zhang, Z., and Yin, X. (2020). A novel probabilistic power flow algorithm based on principal component analysis and high-dimensional model representation techniques. *Energies* 13, 3520. doi:10.3390/en13143520

Liang, H., Hua, H., Qin, Y., Ye, M., Zhang, S., and Cao, J. (2022). Stochastic optimal energy storage management for energy routers via compressive sensing. *IEEE Transactions on Industrial Informatics* 18, 2192–2202. doi:10.1109/tii.2021.3095141. (in press)

Liang, H., Wang, T., Guo, J., and Cao, J. (2021). Probabilistic power flow calculation of microgrid based on ℓ_1-minimization. In *2021 IEEE International Conference on Energy Internet (ICEI)* (IEEE). doi:10.1109/icei52466.2021.00013. (Also available at `http://www.mit.edu/~caoj/pub/doc/jcao_c_powerflow.pdf`)

Lin, X., Jiang, Y., Peng, S., Chen, H., Tang, J., and Li, W. (2020). An efficient nataf transformation based probabilistic power flow for high-dimensional correlated uncertainty sources in operation. *International Journal of Electrical Power & Energy Systems* 116, 105543. doi:10.1016/j.ijepes.2019.105543

Liu, H., Tang, C., Han, J., Li, T., Li, J., and Zhang, K. (2017). Probabilistic load flow analysis of active distribution network adopting improved sequence operation methodology. *IET Generation, Transmission & Distribution* 11, 2147–2153. doi:10.1049/iet-gtd.2016.0406

Liu, Z., Zhang, X., Su, M., Sun, Y., Han, H., and Wang, P. (2020). Convergence analysis of newton-raphson method in feasible power-flow for DC network. *IEEE Transactions on Power Systems* 35, 4100–4103. doi:10.1109/tpwrs.2020.2986706

Ma, H., Wang, H., Xu, X., Yan, Z., and Mao, G. (2021). Efficient probabilistic load flow calculation considering vine copulabased dependence structure of renewable energy generation. *Journal of Donghua University (English Edition)* 38, 6

Marques, E. C., Maciel, N., Naviner, L., Cai, H., and Yang, J. (2019). A review of sparse recovery algorithms. *IEEE Access* 7, 1300–1322. doi:10.1109/access.2018.2886471

Memon, Z. A., Trinchero, R., Manfredi, P., Canavero, F., and Stievano, I. S. (2020). Compressed machine learning models for the uncertainty quantification of power distribution networks. *Energies* 13, 4881. doi:10.3390/en13184881

Papadopoulos, A. D., Zygiridis, T. T., Glytsis, E. N., Kantartzis, N. V., and Antonopoulos, C. S. (2019). Uncertainty study of periodic-grating wideband filters with sparse polynomial-chaos expansions. *IEEE Photonics Technology Letters* 31, 1499–1502. doi:10.1109/lpt.2019.2935322

Rawat, M. S. and Vadhera, S. (2018). Impact of photovoltaic penetration on static voltage stability of distribution networks: A probabilistic approach. *Asian Journal of Water, Environment and Pollution* 15, 51–62. doi:10.3233/AJW-180043

Shen, D., Wu, H., Liu, L., and Gan, D. (2020). Solving probabilistic power flow with wind generation by polynomial chaos expansion method from the perspective of parametric problems. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*

Son, J. and Du, Y. (2021). An efficient polynomial chaos expansion method for uncertainty quantification in dynamic systems. *Applied Mechanics* 2, 460–481. doi:10.3390/applmech2030026

Sun, B., Zhao, W., and Zhu, X. (2016). Compressed sensing for implantable neural recordings using co-sparse analysis model and weighted ℓ_1-optimization

Sun, X., Wang, B., Chen, J., Yinhong, L. I., Zhao, H., Duan, X., et al. (2019). Sparse polynomial chaos expansion based uncertainty quantification for available transfer capability. *Proceedings of the CSEE* 39, 2904–2914. doi:10.13334/j.0258-8013.pcsee.180743. (in Chinese)

Tang, J., Ni, F., Ponci, F., and Monti, A. (2016). Dimension-adaptive sparse grid interpolation for uncertainty quantification in modern power systems: Probabilistic power flow. *IEEE Transactions on Power Systems* 31, 907–919. doi:10.1109/tpwrs.2015.2404841

Wang, S., Du, H., Zhang, G., Lu, J., and Yang, J. (2020). Robust canonical correlation analysis based on l1-norm minimization for feature learning and image recognition. *Journal of Electronic Imaging* 29, 1. doi:10.1117/1.jei.29.2.023001

Wu, H., Zhou, Y., Dong, S., and Song, Y. (2017). Probabilistic load flow based on generalized polynomial chaos. *IEEE Transactions on Power Systems* 32, 820–821. doi:10.1109/tpwrs.2016.2543143